

ANÁLISIS DEL COMPORTAMIENTO DE LA VARIABLE SEXO EN EL LÉXICO DISPONIBLE DE CÁDIZ

ADOLFO GONZÁLEZ MARTÍNEZ
PEPA ORELLANA RAMÍREZ

1. Introducción

Una estratificación clásica en cualquier estudio sociolingüístico es la variable sexo. Desde el trabajo de Sapir 1929 sobre la lengua indígena de California, el Yana, han surgido incesantes ejemplos en los que los contrastes lectales entre hablantes masculinos y femeninos ofrecen todo un conjunto de variantes. Como apunta López Morales 1989, la mayoría de las investigaciones explican que las diferencias entre hablantes masculinos y femeninos son debidas al factor prestigio. *Los datos acumulados hasta la fecha demuestran de manera evidente que las mujeres son más conscientes de la valoración que su comunidad hace de los fenómenos del lenguaje, y apoyan aquellos que obtienen más alto estatus en la evaluación social. (...) El habla de las mujeres no sólo es diferente al habla de los hombres sino que es mejor socialmente hablando. Así como se espera que la conducta social de las mujeres sea más correcta, también su habla debe serlo* (López Morales, 1989:125-126).

El objetivo de este trabajo es comprobar si estas diferencias se producen igualmente en el léxico de la provincia de Cádiz, para lo cual hemos acudido al repertorio de términos obtenido en la investigación del léxico disponible de la provincia de Cádiz (González Martínez, 1997), pues creemos que es un inventario de vocablos muy válido para analizar el comportamiento léxico de los hablantes. El léxico disponible *es el conjunto de vocablos que utiliza una comunidad de hablantes como parte medular de su léxico patrimonial pero que únicamente aparecen en situaciones precisas, por lo que no son muy frecuentes pero sí muy conocidos. Son inestables debido a que es necesario que se toque un tema específico en el discurso para que aparezcan* (López Chávez, 1994:69). La disponibilidad, por lo tanto, agrupa el vocabulario que el hablante podría utilizar en un contexto comunicativo dado (Sánchez Morales y Murillo

Rojas, 1993). Toda lengua posee un amplio conjunto de palabras de contenido semántico muy concreto que los hablantes únicamente manejan si lo permite el tema del discurso; en el lexicón mental existe una serie de términos que no se actualizan a menos que se necesiten para comunicar una información muy precisa. Se trata, pues, de un léxico disponible cuyo estudio no puede emprenderse manejando frecuencias, ya que este factor es pertinente en el caso de las actualizaciones léxicas efectivas, no potenciales (Aitchison 1987).

Lorán y López Morales 1983 mencionan la necesidad de establecer un orden en las palabras producidas por los informantes en las pruebas asociativas escritas que considere, además de la frecuencia, la jerarquización dada por los sujetos encuestados.

En ese mismo año comienzan a diseñarse fórmulas capaces de ponderar la frecuencia de las palabras con los lugares que estas ocuparan en las listas, de ese año es la fórmula Lorán-López Morales 1983 y poco tiempo después, la de López Chávez-Strassburger 1991, que partían ambas de la misma base: desarrollar una fórmula matemática que fuera capaz de ponderar adecuadamente la frecuencia adquirida por una lexía con sus lugares de aparición en las listas. No hay que olvidar que aquellas palabras que acuden primero a la memoria como reacción a un estímulo determinado son realmente las más disponibles.

La segunda fórmula fue la que se aplicó en la investigación del Léxico Disponible de Cádiz, porque aunque la primera dio excelentes resultados en los casos de listas cerradas, las primeras veinte palabras, por ejemplo, no fue así en las listas abiertas, puesto que pierde su poder discriminador a partir del vigésimo tercer lugar y las encuestas de dicho estudio están unificadas por el tiempo, no por la cantidad de lexías de las respuestas.

Del análisis de una matriz de vectores se ha desprendido, después de una reiterada experimentación, la fórmula siguiente¹:

$$D(P_j) = \sum_{i=1}^n e^{-2.3 \frac{(i-1)}{(n-1)}} \times \frac{f_{ji}}{I_i}$$

donde

$D(P_j)$ = disponibilidad de la palabra j .

n = máxima posición alcanzada en el centro de interés.

i = número de posición de que se trata.

j = índice de la palabra en cuestión.

e = número natural.

f_{ji} = frecuencia absoluta de la palabra j en la posición i .

I_i = número de informantes que participaron en la prueba.

¹ Una explicación más amplia y sistemática sobre las fórmulas para obtener los diferentes índices de disponibilidad léxica puede encontrarse en Butrón 1991.

López Chávez y Strassburger 1991 afirman que los valores de la matriz que deberían incluir por su relevancia lingüística dentro de la teoría de la disponibilidad eran:

- a) La frecuencia absoluta con que fue dicha cada palabra en cada posición (f_{ji}).
- b) La frecuencia absoluta de la palabra, que resulta de sumar las diferentes frecuencias en cada posición.
- c) El número de informantes participantes en la prueba (I_j).
- d) El número de posiciones alcanzadas en la encuesta en el centro de interés (n).
- e) Las posiciones en que fue dicha cada palabra (i).

La aplicación del número e elevado al exponente es el verdadero ponderador de la posición, lo que permite combinar, sin distorsión, la frecuencia y la posición de emisión de cada palabra, ya que arroja una ponderación variable entre el 1 y 0,1, independientemente del número de participantes, de la extensión de los listados producidos por cada informante, del número de sujetos que llegan a cada posición y de la frecuencia de aparición del vocablo (López Chávez, 1992a y b).

2. Muestra

La investigación se ha realizado sobre una muestra empírica representativa integrada por veinte grupos de COU de otros tantos Centros Educativos de Cádiz, distribuidos por toda la provincia, siendo el total de los encuestados cuatrocientos, veinte de cada instituto. Para recaudar este léxico, disponible pero no frecuente, es necesaria la aplicación de una prueba asociativa en torno a unos estímulos, centros de interés, en nuestro caso los dieciséis comunes a todas las investigaciones que sobre el Léxico Disponible se están llevando a cabo tanto en España como en Hispanoamérica, de tal manera que puedan servir de base para ulteriores comparaciones dialectales. Los centros de interés que se les presentaron a los encuestados como estímulos fueron: 1. Partes del cuerpo. 2. La ropa. 3. Partes de la casa (sin los muebles). 4. Los muebles de la casa. 5. Alimentos y bebidas. 6. Objetos colocados en la mesa para la comida. 7. La cocina y sus utensilios. 8. La escuela: muebles y materiales. 9. Iluminación, calefacción y medios para airear un recinto. 10. La ciudad. 11. El campo. 12. Medios de transporte. 13. Trabajos del campo y del jardín. 14. Los animales. 15. Juegos y distracciones. 16. Profesiones y oficios.

La investigación llevada a cabo en la provincia de Cádiz analiza los resultados obtenidos por la totalidad de los encuestados y por distintas variables, entre las que se encuentra el sexo, que es en la que centramos el estudio en el presente artículo.

La distribución por sexo de los informantes de nuestro trabajo la podemos observar en el siguiente cuadro:

	UNIVERSO	MUESTRA
HOMBRES	560.379	181
MUJERES	567.243	219
TOTAL	1.127.691	400

Cuadro 1. Distribución por sexo

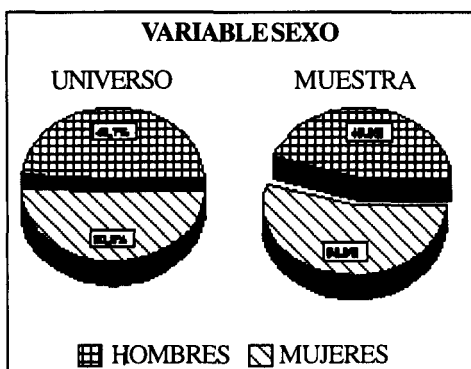


Gráfico 1. Distribución por sexo

3. Análisis cuantitativo

La totalidad de los informantes produjo 118.777 unidades léxicas, lo que supone una media de 296,94 palabras por individuo encuestado. Las mujeres superan a los hombres en seis palabras de media (300 frente a 294) lo que no es una diferencia muy amplia pero sí significativa. Estos promedios no son fácilmente alterables por unos pocos valores individuales atípicos, por lo que nos parece una buena forma de medir la riqueza léxica de un centro de interés o de un grupo de sujetos en particular (Echeverría, 1991). Las mujeres sólo son aventajadas por los hombres en los centros *1 partes del cuerpo*, *9 iluminación, calefacción* y *medios para airear un recinto*, *10 la ciudad*, *11 el campo*, *12 medios de transporte* y *13 trabajos del campo y del jardín*; coinciden en el centro *3 partes de la casa (sin los muebles)* y dominan a los informantes masculinos en los restantes centros de interés. Por lo tanto podemos deducir, con todas las reservas que queramos, que las mujeres poseen una mayor riqueza léxica ya que emiten un mayor número de palabras por sujeto.

Con relación al número de vocablos, palabras distintas, puede parecer que una mayor cantidad de vocablos en un determinado centro de interés es indicación de una superior riqueza léxica en el mismo al compararlo con otro de menor número de vocablos; sin embargo, no es así necesariamente. Este índice sería un buen indicador sólo si el grupo fuera muy homogéneo, pero basta que haya uno o dos individuos atípicos en el grupo para que el índice suba desproporcionadamente (Echeverría, 1991). Por ello nos parece conveniente manejar este indicador con más reservas que el anteriormente analizado de la media de palabras.

La totalidad de los informantes produjo 6.655 vocablos, los hombres aportaron 5.163 y las mujeres 4.813, superando a los hombres en el centro 3 *los muebles de la casa* y en el 7 *la cocina y sus utensilios*, lo que nos hace pensar que las mujeres, a pesar de los cambios acaecidos en la sociedad en cuanto a la liberación femenina, dominan en dos centros de interés de temática tradicionalmente femenina, puesto que uno y otro centros se refieren al hogar, competencia de la mujer desde épocas pretéritas.

Por último, en este apartado estudiaremos la densidad de la disponibilidad; este índice describe la naturaleza cuantitativa de cada centro de interés independientemente y propicia el establecimiento de comparaciones entre los centros del corpus; consiste en poner en relación la cantidad de palabras que ofrece el corpus para cada centro de interés con el total de vocablos de ese mismo centro. Con ello sabemos qué centro ofrece más dispersión (aquel cuyo índice de densidad sea más bajo) y qué centro ofrece más homogeneidad (aquel cuyo índice sea más alto). Cuanto más variadas sean las respuestas, menor será el índice de densidad, es decir, la dispersión será mayor. De este modo, el índice de densidad de la disponibilidad informa de si el centro de interés es cerrado o compacto, o bien, abierto o difuso.

En los resultados totales, el centro de interés 4 *los muebles de la casa* es el que presenta mayor cohesión, es decir, es el más compacto, mientras que el número 11 *el campo* es el más disperso, porque los informantes lo abordan desde perspectivas distintas: el campo como agricultura, como ecología, como hábitat y como lugar de deporte y esparcimiento.

Tanto los hablantes masculinos como las informantes femeninas coinciden en los centros más dispersos, 11 *el campo* y 15 *juegos y distracciones*; el centro que más cohesión presenta es el 2 *la ropa*, tanto en los hombres como en las mujeres; por ello podríamos concluir afirmando que, en cuanto a la densidad de la disponibilidad, no hay diferencias debidas a la variable sexo.

4. Análisis cualitativo

En el presente apartado analizaremos la naturaleza del léxico obtenido en las encuestas y comprobaremos que las diferencias entre los informantes masculinos

nos y los femeninos son interesantes cuando nos centramos en el estudio de los términos coloquiales utilizados por unos y por otras.

Para determinar los términos coloquiales hemos acudido a una bibliografía especializada², aunque siempre teniendo como referencia el *DRAE*³. En este apartado englobamos los tradicionalmente denominados vulgarismos (Lapesa, 1980), las palabras malsonantes, los términos jergales y los socialmente estigmatizados, es decir, todos aquellos que pertenecen al registro menos cuidado del habla. Para realizar este análisis hemos elegido, por razones de espacio, únicamente tres centros de interés de los que componen la encuesta de Disponibilidad Léxica: el 1 *partes del cuerpo*, el 10 *la ciudad* y el 15 *juegos y distracciones*, ya que en ellos es donde aparece el mayor número de coloquialismos de las listas.

En el centro de interés 1 *partes del cuerpo* los hombres registran diecinueve términos coloquiales mientras que las mujeres anotan diez, pero además de la estimable diferencia cuantitativa se aprecia también que en las nóminas de informantes masculinos la mayoría de los términos aparece en una posición más adelantada que en las de las informantes femeninas. Lo observamos más esquemáticamente en el siguiente cuadro, en el que aparece, junto con el vocablo, el puesto que ocupa este en las listas. Asimismo marcamos en negrita únicamente aquellos vocablos que aparecen en un lugar superior en el repertorio de las mujeres o los que solamente aparecen en estas:

VOCABLOS	Hombres	Mujeres
CULO	21	19
TETA	471	60
PICHA	591	89
SOBACO	74	104
COÑO	77	107
HUEVO	91	Ø
CHOCHO	101	114
NABO	112	Ø
COJÓN	145	Ø
PINREL	175	Ø

² Cela, C.J. 1969, 1971; González Martínez 1986; León 1980; Martín 1979 y Oliver 1985.

³ Nos referimos a la 21ª edición de 1992, en su versión CD ROM.

ANÁLISIS DEL COMPORTAMIENTO DE LA VARIABLE SEXO EN EL LÉXICO DISPONIBLE

CHUMINO	201	Ø
COLA	203	Ø
CAPULLO	208	Ø
PITO	211	Ø
CARAJÓ	219	Ø
CACHA	221	124
CONEJO	223	Ø
ALMEJA	226	Ø
POLLA	252	207
TOTO	Ø	121
TETE	Ø	216

Cuadro 2. Coloquialismos en el centro 1

Si analizamos el centro de interés 10 *la ciudad* advertimos que los hombres registran once términos coloquiales y las mujeres únicamente dos; en este estímulo no podemos hablar de diferencia de posición de los vocablos en las nóminas, sino de resultados dispares entre los informantes masculinos y las informantes femeninas.

En el siguiente cuadro vemos más esquemáticamente las diferencias:

VOCABLOS	H	M
PUTICLUB	103	511
MIERDA	197	Ø
CAMELLO	239	Ø
TÍA	279	Ø
CONDÓN	351	Ø
PUTA	370	Ø
PORRETA	382	Ø
CHORIZO	390	Ø
MADERO	469	Ø
MOJÓN	488	Ø
CHAPERO	551	Ø
DROGATA	Ø	286

Cuadro 3. Coloquialismos en el centro 10

Por último, vamos a analizar el centro de interés *15 juegos y distracciones*, en el que los informantes masculinos aportan trece vocablos, mientras que las informantes femeninas únicamente tres; algunos de los términos no aparecen en las listas de las mujeres por razones obvias, ya que se trata de comportamientos biológicos masculinos. En el siguiente cuadro los observamos:

VOCABLOS	H	M
TETO ⁴	54	Ø
FOLLAR	59	183
ENROLLARSE	155	242
SESENTA Y NUEVE	163	Ø
HACERSE UNA PAJA	166	Ø
PORRO	207	356
JODER	243	Ø
DAR POR CULO	277	Ø
COGER EL CULO	294	Ø
PUTICLUB	350	Ø
PIBA	357	Ø
TRIFI	365	Ø
ECHAR UN POLVO	416	Ø

Cuadro 4. Coloquialismos en el centro 15

5. Conclusiones

La primera conclusión que podemos extraer es que en el Léxico Disponible de la provincia de Cádiz las diferencias debidas a la variable sexo se concentran fundamentalmente en la naturaleza del léxico, básicamente en el uso, por parte de las hablantes femeninas, de un léxico más prestigiado socialmente y a la vez más conservador (características que suelen ir juntas en numerosos casos), lo cual viene a corroborar nuestra hipótesis de partida acerca de la índole de la discrepancia léxica entre hombres y mujeres. Es evidente que la igualdad entre los sexos, al menos desde el punto de vista jurídico, está prácticamente conseguida, sin embargo resulta llamativo que en una encuesta anónima, realizada a jóvenes de COU, el léxico femenino siga manifestando las características tradicionales de escasez de términos socialmente estigmatizados, lo cual nos lleva a concluir que la lengua avanza a un ritmo más lento que la sociedad que la utiliza.

⁴ Responde a un juego de palabras en el que el chico le pregunta a la chica «¿juegas al teto?» y la chica debe responder «¿qué es el teto?» a lo que el chico dice: «tú te agachas y yo te la meto».

También destacamos que las mujeres aventajan a los hombres en media de palabras por sujeto, lo cual nos conduce a señalar que poseen una mayor riqueza léxica que los varones, aunque no sea excesivamente manifiesta.

Una tercera conjetura sería la permanencia de factores sociales tradicionales en las informantes femeninas, lo que se deduce de su mayor aportación de vocablos en aquellos centros cuya temática podríamos considerar clásicamente femenina, como pueden ser los muebles de la casa y la cocina y sus utensilios. En resumen, en este estudio de la variable sexo en la provincia de Cádiz, queda confirmado que existen diferencias léxicas entre hablantes masculinos y femeninos, y que esas diferencias se deben fundamentalmente a cuestiones de prestigio, como decíamos al comenzar este artículo.

BIBLIOGRAFÍA

- AITCHISON, Jean, 1987, *Words in the mind: An introduction to the mental lexicon*, Oxford-New York, Basil Blackwell Ltd.
- BUTRÓN, Gloria, 1991, *Nuevos índices de disponibilidad léxica*, López Morales, ed.
- CELA, Camilo José, 1969, *Diccionario Secreto, I*, Madrid, Alfaguara.
- , 1971, *Diccionario Secreto, II*, Madrid, Alfaguara.
- ECHEVARRÍA, Max S., 1991, *Crecimiento de la disponibilidad léxica en estudiantes chilenos de nivel básico y medio*, López Morales, ed.
- GONZÁLEZ MARTÍNEZ, Adolfo, 1986, «Eufemismos de la esfera sexual en las coplas del carnaval de Cádiz». En *Boletín de la Academia Puertorriqueña de la Lengua Española*. XIV, San Juan.
- , 1997, «La Disponibilidad Léxica de Cádiz», tesis doctoral inédita. Universidad de Cádiz.
- LAPESA, Rafael, 1980, *Historia de la Lengua Española*, Madrid, Gredos, (8ª ed.).
- LEÓN, Víctor, 1980, *Diccionario de argot español*, Madrid, Alianza.
- LÓPEZ CHÁVES, Juan, 1992a, «Léxico fundamental panhispánico: realidad o utopía». En *Actas del IV Congreso Internacional sobre el español de América*, en prensa.
- , 1992b, «Alcances panhispánicos del léxico disponible». En *Linguística*, 4: 26-124.
- , 1994, «Comportamiento sintáctico de algunos verbos ordenados según su grado de disponibilidad léxica». En *REALE*, 1, Universidad de Alcalá, Alcalá de Henares.
- , y STRASSBURGER, Carlos, 1991, *Un modelo para el cálculo de disponibilidad léxica individual*. En López Morales, ed.
- LÓPEZ MORALES, Humberto, 1989, *Sociolingüística*, Madrid, Gredos.
- LORÁN, Roberto y LÓPEZ MORALES, Humberto, 1983, *Nouveau calcul de l'indice de disponibilité*, MS.
- MARTÍN, Jaime, 1979, *Diccionario de expresiones malsonantes del español*, Madrid, ISTMO, (2ª ed.).
- OLIVER, J. M., 1985, *Diccionario de argot*, Madrid, Sena.
- REAL ACADEMIA ESPAÑOLA, 1992, *Diccionario de la lengua española*, Madrid, Espasa Calpe, 21ª ed.
- SÁNCHEZ MORALES, Víctor y Marielos MURILLO ROJAS, 1993, «Campos semánticos y disponibilidad léxica en preescolares». En *Revista Educación*, 17₃, Costa Rica.
- SAPIR, Edward, 1929, *Male and female forms of speech in Yana*, reimpresso en los *Selected Writings of Edward Sapir*, ed. Por D.G. Mandelbaum, Berkeley, 1949.